猴樟叶绿体基因组特征分析

赵渊祥[1,2]，梁大曲[1,2]，谢双琴[1,2]，王好运[1,2]，吴峰[1,2*]

（1. 贵州大学 贵州省森林资源与环境研究中心，贵州省高原山地林木培育重点实验室，贵阳 550025；2. 贵州大学 林学院，贵阳 550025）

摘 要：猴樟（*Cinnamomum bodinieri*）枝叶含有丰富的精油，是重要的园林绿化树种和经济树种，但目前有关猴樟基因组学的研究报道不多。为揭示猴樟叶绿体基因组特征及系统发育关系，该文基于高通量测序平台进行测序，从头组装了完整的猴樟叶绿体基因组，并对其基因组结构、基因构成及序列重复、密码子使用偏好性进行分析，结合樟亚科主要属物种叶绿体基因组数据构建系统发育树。结果表明：（1）猴樟叶绿体基因组全长 152 727 bp，包括一对 20 132 bp 的反向重复序列（IRs）区、93 605 bp 的大单拷贝（LSC）区和 18 858 bp 的小单拷贝（SSC）区，总 GC 含量为 39.13%。（2）该基因组共编码 127 个基因，包括 83 个蛋白质编码基因（PCGs）、36 个转运 RNA 基因（tRNAs）和 8 个核糖体 RNA 基因（rRNAs）；共鉴定出 92 个 SSR 位点，其中大部分是 A/T 组成的单核苷酸重复序列；密码子适应指数（CAI）为 0.166，有效密码子数（ENc）为 54.68；猴樟与近缘种的叶绿体基因组主要在 IR 区和两个 SC 区边界上存在一定的差异。（3）24 种樟亚科植物的系统发育树显示，猴樟与樟树亲缘关系最近，同时支持了樟属-甜樟属分支(*Cinnamomum*-*Ocotea* Clade)、月桂属-新木姜子属分支(*Laurus*-*Neolitsea* Clade)、润楠属-鳄梨属分支(*Machilus*-*Persea* Clade)的建立。该研究丰富了猴樟遗传资源信息，进一步确定了樟亚科主要属的系统发育地位。

关键词：猴樟，叶绿体基因组，SSR，密码子使用偏好性，系统发育

中图分类号：Q943　　　　**文献标识码**：A

Analysis on chloroplast genomic characteristics of *Cinnamomum bodinieri*

ZHAO Yuanxiang[1,2], LIANG Daqu[1,2], XIE Shuangqin[1,2], WANG Haoyun[1,2], WU Feng[1,2*]

（1. *Institute for Forest Resources and Environment of Guizhou*, *Key Laboratory of Forest Cultivation in Plateau Mountain of Guizhou Province*, *Guizhou University*, Guiyang 550025, China; 2. *College of Forest*, *Guizhou University*, Guiyang 550025, China）

Abstract: *Cinnamomum bodinieri* is important landscaping and economic tree species, which is rich in essential oils in branches and leaves. However, there are few theoretical researches on the genomics of *C. bodinieri*. In order to reveal the chloroplast genomic characteristics and phylogenetic relationship of *C. bodinieri*, the complete chloroplast genome was sequenced based on Illumina platform and assembled through de novo. The genome structure, gene composition, repeats, codon usage bias and phylogeny were analyzed subsequently. Furthermore, the phylogenetic tree was constructed with the chloroplast genome data of the main species of Subfam. Lauroideae. The results were as follows: (1) The complete chloroplast genome of *C. bodinieri* was 152 727 bp in length including two inverted repeats (IRs) of 20 132 bp, which were separated by large single copy(LSC) and short single copy(SSC) of 93 605 bp and 18 858 bp, respectively, and the GC content was 39.13%. (2) The genome encoded 127 functional genes, including 83 protein-coding genes, 36 tRNA genes, and 8 rRNA genes. 92 SSR loci were detected in the chloroplast genome, and most of the SSRs were composed of nucleobase A and T. The codon adaptation index and

收稿日期：2022-08-05

基金项目：贵州省种苗站项目（2022-ZM-05）；贵州省科技厅计划项目（黔科合平台人才[2018]5261）；贵州省一流学科建设项目（GNYL[2017]007）[Supported by Guizhou Province Seedling Station Project (2022-ZM-05); Guizhou Provincial Department of Science and Technology Planning Project (QKHPTRC[2018]5261); First-class Discipline Construction Project in Guizhou Province (GNYL[2017]007)]。

第一作者：赵渊祥(1998-)，硕士研究生，主要从事林木遗传育种研究，(E-mail) gs.yxzhao20@gzu.edu.cn。

*__通信作者__：吴峰，博士，教授，主要从事林木遗传育种研究，(E-mail) fwu@gzu.edu.cn。

effective codon number were 0.166 and 54.68, respectively. There were some differences in IR region and the boundary of two SC regions of the chloroplast genomes between *C. bodinieri* and related species. (3) Phylogenetic tree based on 24 species of Subfam. Lauroideae showed that the *C. bodinieri* was closely related to *C. camphora*. The phylogeny strongly supported the establishment of the three clades, *Cinnamomum-Ocotea*, *Laurus-Neolitsea*, and *Machilus-Persea*. This study enriched the information on the genetic resources of *C. bodinieri*, and further clarified the phylogenetic status of the main genera of Subfam. Lauroideae.

Keyword: *Cinnamomum bodinieri*, chloroplast genome, SSR, codon usage bias, phylogenetic

叶绿体是植物细胞特有的细胞器，也是光合作用的主要场所，它们拥有独立于核基因组的完整的叶绿体基因组（Shinozaki et al., 1986）。大多数陆生植物叶绿体基因组大小在 107～218 kb 之间，与核基因组和线粒体组相比，遗传信息的携带量相对较少（Daniell et al., 2016）。但叶绿体基因组依赖于母系遗传，具有易于提取和纯化、遗传信息丰富、共线性良好、高度保守的序列和大量的简单序列重复（Simple sequence repetition, SSR）基因座等优点（Dobrogojski et al., 2020）。近年来随着高通量测序技术的发展，研究者已相继组装、破译近千个物种的叶绿体基因组，揭示了植物物种内部和物种之间在序列和结构方面的变异，这些研究对一些植物的系统发育，特别是阐明进化分支内的物种进化关系做出了重大的贡献（Njuguna et al., 2013; Daniell et al., 2016）。

樟科植物广泛分布于世界各地的热带与亚热带地区，包含 50 多个属 3 500 多种（Chanderbali et al., 2001）。早期分类系统中，研究者利用形态学特征将樟科分为樟亚科和无根藤亚科（Kostermans, 1957），而一直以来樟亚科的系统发育关系存在争议，在是否建立樟族等族群以及檫木属、新樟属等的系统发育地位等问题分歧较多。随后基于单个或多个叶绿体关键基因区段等分子序列数据构建的系统发育树大大提高对樟亚科植物分类的理解，但依然存在一个分辨率较低的末端分支，包含樟族、月桂族等，难以找到形态学上的共有衍征来阐明系统发育关系（田永靖，2021）。近年来研究者基于叶绿体基因组将樟亚科分为 *Hypodaphnis* 分支、*Beilschmiedia-Cryptocarya* 分支、*Neocinnamomum* 分支、*Caryodaphnopsis* 分支、*Chlorocardium-Mezilaurus* 分支、*Machilus-Persea* 分支、*Cinnamomum-Ocotea* 分支和 *Laurus-Neolitsea* 分支（Song et al., 2020），其中 *Machilus-Persea* 分支、*Cinnamomum-Ocotea* 分支和 *Laurus-Neolitsea* 分支是樟亚科植物分类系统中争议较多的类群。已有研究表明，叶绿体基因组在纠正物种错误鉴定以及发现隐存种与新物种方面具有一定优势（Liu et al., 2021）。因此，在目前樟科植物核基因组、线粒体基因组的测序、组装技术暂未普及的前提下，利用叶绿体基因组构建樟科系统发育树仍然是当前的有效手段之一（Song et al., 2020; Liu et al., 2021）。

猴樟（*Cinnamomum bodinieri*）是樟科樟属常绿阔叶树，原产于中国，主要分布在贵州省、湖南省西部和湖北省西部（Fang et al., 2011）。猴樟外形美观，树冠厚实，木材质地坚硬，有光泽，有香味，枝叶含有丰富的精油（Xiao et al., 2020），是重要的园林绿化树种和经济树种。已有研究主要集中在其生态特征（田小琴和韦小丽，2011）、抗逆性（宋芳琳等，2012）、栽培和育种（张怡等，2014），组织培养技术（肖祖飞等，2020）等方面，鲜有关于猴樟叶绿体基因组及系统发育方面的研究。本研究基于 Illumina 测序平台，对贵州产地猴樟的叶绿体全基因组进行测序、组装和注释，从叶绿体基因组结构、SSR 位点、密码子使用偏性、IR 区的扩张与收缩以及系统发育等方面开展研究，同时结合樟亚科主要属的 23 个物种构建系统进化树，拟揭示猴樟叶绿体基因组以下信息：（1）猴樟叶绿体基因组的基本特征；（2）猴樟叶绿体基因组 SSR 位点及基因密码子偏好性情况；（3）猴樟与近缘种的叶绿体基因组 IR 区结构差异；（4）猴樟叶绿体基因组系统发育所属分支。

1 材料与方法

1.1 试验材料

猴樟叶片采自贵州省贵阳市贵州大学南校区（106°67′ E，26°43′ N），经鉴定为樟科樟属猴樟。样本经液氮速冻处理后保存于-80 ℃冰箱。

1.2 叶绿体全基因组 DNA 提取与测序

选取叶片组织，利用植物基因组 DNA 提取试剂盒（TIANGEN Beijing China）提取猴樟 DNA，经过 1%琼脂糖凝胶电泳和紫外分光光度计检测质量后利用 Illumina NovaSeq 平台完成高通量测序。具体实验流程按照 Illumina 公司提供的标准方法执行，将 DNA 片段化后，对片段化的 DNA 进行片段纯化、末端修复、3′ 端加 A、连接测序接头，经 PCR 扩增形成测序文库，文库质检合格后进行测序。

1.3 叶绿体全基因组的组装与注释

采用 SPAdes（v3.10.1）软件（Bankevich et al., 2012）组装叶绿体基因组。采用两种方法对叶绿体基因组进行注释，以提高注释的准确性。首先，使用 prodigal（v2.6.3）软件（Hyatt et al., 2010）注释叶绿体的 CDS，使用 hmmer（v3.1b2）软件（Collyda et al., 2006）预测 rRNA，使用 aragorn（v.1.2.38）软件（Laslett & Canback, 2004）预测 tRNA。其次，根据 NCBI 上已经公布的近缘物种，提取其基因序列，然后使用 blast（https://blast.ncbi.nlm.nih.gov/Blast.cgi）比对组装的序列，得到第二种注释结果。然后通过对比去除错误注释及冗余的注释，确定多外显子边界，获得最终的注释。最后将注释完成的猴樟叶绿体基因组序列提交至 NCBI（https://www.ncbi.nlm.nih.gov/），获得登录号（MW381013）。

1.4 叶绿体基因组 SSR 位点分析

使用 MISA（v1.0）软件（Beier et al., 2017）搜索叶绿体基因组中的 SSR 标记，设置单核苷酸重复次数 ＞10，二核苷酸重复次数 ＞5，三核苷酸重复次数 ＞4，四核苷酸、五核苷酸和六核苷酸 SSR 重复次数 ＞3。

1.5 密码子偏好性分析

根据 127 个基因的 CDS 序列，筛选唯一的（多个拷贝基因选择一个拷贝）且序列长度大于 300 bp 的 CDS，并用 CodonW（v.1.4.2）软件（Wong et al., 2010）估计每个密码子的相对同义密码子使用（relative synonymous codon usage, RSCU）、有效密码子数（effective number of codons, ENc）和密码子适应指数（codon adaptation index, CAI）。

1.6 边界分析

利用近缘物种新樟（*Neocinnamomum delavayi*）、肉桂（*Cinnamomum cassia*）、檫木（*Sassafras tzumu*）、月桂（*Laurus nobilis*）和大叶新木姜子（*Neolitsea levinei*）叶绿体基因组，通过 Irscope（https://irscope.shinyapps.io/irapp/）在线软件（Amiryousefi et al., 2018）进行可视化对比，分析 4 个区域的变化，特别是 IR 区的扩张和收缩以及 IR/SC 边界基因种类和位置的变化。

1.7 叶绿体基因组系统发育分析

利用猴樟叶绿体基因组和从 NCBI 下载的 23 个樟亚科（Subfam. Lauroideae）植物叶绿体基因组构建系统发育树。利用 MAFFT（v.7.475）软件（Nakamura et al., 2018）进行多序列比对，然后利用 IQ-TREE（v. 2.0.3）软件（Nguyen et al., 2015）构建系统发育树，步长自检 1 000 次。

2 结果与分析

2.1 叶绿体基因组结构

猴樟叶绿体基因组全长为 152 727 bp（图 1），呈典型的四分体结构，由 93 605 bp 的大单拷贝区（large single copy, LSC）、18 858 bp 的小单拷贝区（short single copy, SSC）和 20 132 bp 的一对反向重复区（inverted repeats, IRs）组成。基因组总 GC 含量为 39.13%，IR 区 GC 含量最高（44.37%），其次为 LSC（37.94%），SSC 最低（33.83%）。叶绿体基因组共有 127 个基因，包括蛋白质编码基因 83 个、tRNA 基因 36 个、rRNA 基因 8 个。根据功能将基因分成 4 类（表 1），其中光合作用相关基因 45 个，自我复制相关基因 70 个，其他蛋白编码基因 5 个，未知功能基因 7 个。猴樟的叶绿体基因组中，有 1 个多拷贝基因（*tRNA*-Leu）、4 个三拷贝基因（*tRNA*-Ser、*tRNA*-Arg、*tRNA*-Met 和 *tRNA*-Val）和 12 个双拷贝基因；17 个基因含有内含子，其中 14 个基因含有 1 个内含子、3 个基因（*ycf*3、*clp*P 和 *rps*12）含有 2 个内含子。

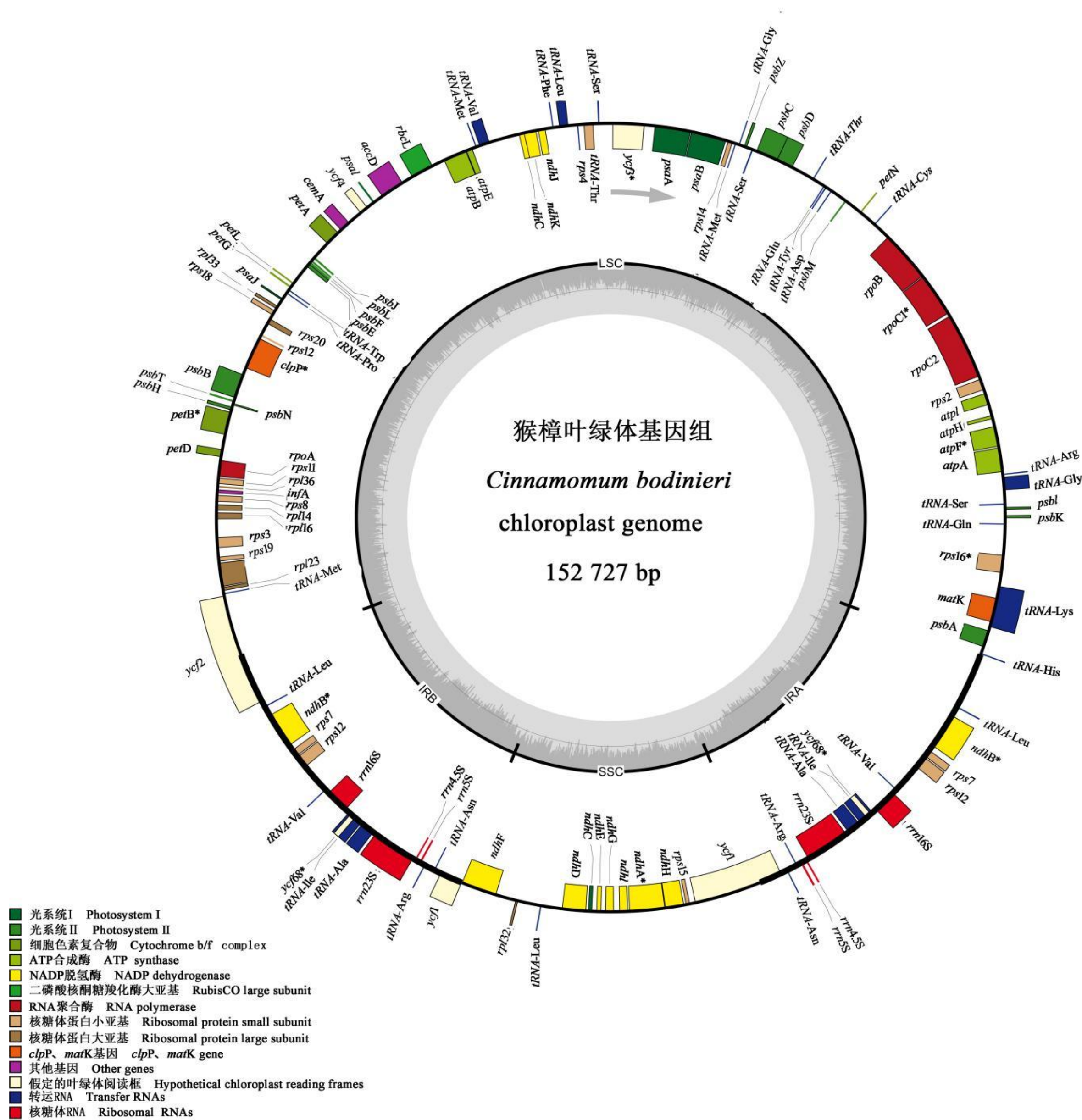


图 1 猴樟叶绿体基因组图谱

Fig.1 Chloroplast genome map of *Cinnamomum bodinieri*

表 1 猴樟叶绿体基因组注释基因信息

Table 1 Annotated gene information of *Cinnamomum bodinieri* chloroplast genome

基因功能分类 Gene functional classification	基因种类 Gene group	基因 ID Gene ID
光合作用相关基因 Genes related to photosynthesis	光系统I Photosystem I	*psa*B，*psa*A，*psa*I，*psa*J，*psa*C
	光系统II Photosystem II	*psb*A，*psb*K，*psb*I，*psb*M，*psb*D，*psb*C，*psb*Z，*psb*J，*psb*L，*psb*F，*psb*E，*psb*B，*psb*T，*psb*N，*psb*H

	细胞色素 b/f 复合体 Cytochrome b/f complex	*pet*N，*pet*A，*pet*L，*pet*G，*pet*B^{a}，*pet*D
	ATP 合酶 ATP synthase	*atp*A，*atp*F^{a}，*atp*H，*atp*I，*atp*E，*atp*B
	NADH 脱氢酶 NADH dehydrogenase	*ndh*J，*ndh*K，*ndh*C，*ndh*B^{*2a}，*ndh*F，*ndh*D，*ndh*E，*ndh*G，*ndh*I，*ndh*A^{a}，*ndh*H
	二磷酸核酮糖羧化酶大亚基 RubisCO large subunit	*rbc*L
自我复制相关基因 Self-replicating related genes	RNA 聚合酶 RNA polymerase	*rpo*C2，*rpo*C1^{a}，*rpo*B，*rpo*A
	核糖体小亚基 Ribosomal protein small subunit	*rps*12^{*2b}，*rps*16^{a}，*rps*2，*rps*14，*rps*4，*rps*18，*rps*11，*rps*8，*rps*3，*rps*19，*rps*7^{*2}，*rps*15，*rps*7
	核糖体大亚基 Ribosomal protein large subunit	*rpl*33，*rpl*20，*rpl*36，*rpl*14，*rpl*16，*rpl*2^{a}，*rpl*23，*rpl*32
	转运 RNA Transfer RNAs	*tRNA*-His，*tRNA*-Lys^{a}，*tRNA*-Gln，*tRNA*-Ser^{*3}，*tRNA*-Gly^{*2a}，*tRNA*-Arg^{*3}，*tRNA*-Cys，*tRNA*-Asp，*tRNA*-Tyr，*tRNA*-Glu，*tRNA*-Thr^{*2}，*tRNA*-Leu^{*4a}，*tRNA*-Phe，*tRNA*-Met^{*3}，*tRNA*-Trp，*tRNA*-Pro，*tRNA*-Val^{*3}，*tRNA*-Ala^{a}，*tRNA*-Arg，*tRNA*-Asn^{*2}，*tRNA*-Ile^{a}，*tRNA*-Val^{a}
	核糖体 RNA Ribosomal RNAs	*rrn*16S^{*2}，*rrn*4.5S^{*2}，*rrn*5S^{*2}，*rrn*23S^{*2}
其他基因 Other genes	成熟酶基因 Maturase gene	*mat*K
	翻译起始因子 Translational initiation factor	*inf*A
	乙酰辅酶 A 羧化酶亚基 Subunits of Acetyl-CoAcarboxylase	*acc*D
	包膜蛋白基因 Envelop membrane protein gene	*cem*A
	蛋白酶基因 Protease gene	*clp*P^{b}
未知功能基因 Unknown genes	保守开放性阅读框 Conserved open reading frames	*ycf*3^{b}，*ycf*4，*ycf*2，*ycf*1^{*2}，*ycf*68^{*2a}

注: ^{*2}和^{*3}分别表示该基因拷贝 2，3 次并以此类推; a 和 b 各自表示有 1 和 2 个内含子。

Note: ^{*2} and ^{*3} indicate that the gene is copied 2, 3 times respectively and so on; a and b indicate that there are 1 and 2 introns, respectively.

2.2 叶绿体基因组 SSR 位点分析

猴樟叶绿体基因组中共检测到 92 个 SSR 位点（图 2）。SSR 主要分布在 LSC 区（70 个，76.1%），其次是 SSC 区（18 个 SSR）和 IR 区（4 个 SSR）。SSR 位点中包括 68 个单碱基重复序列（73.9%）、

10 个二碱基重复序列（10.9%）、3 个三碱基重复序列（3.3%）、8 个四碱基重复序列（8.7%）、2 个五碱基重复序列（2.2%）以及 1 个六碱基重复序列（1.1%）。单核苷酸 SSR 为多聚腺嘌呤（A）和多胸腺嘧啶（T）重复。SSR 基因序列碱基组成中，A 或 T 组成的 SSR 占 85.8%，表明猴樟叶绿体的 SSR 中碱基 A 和 T 有明显的优势。

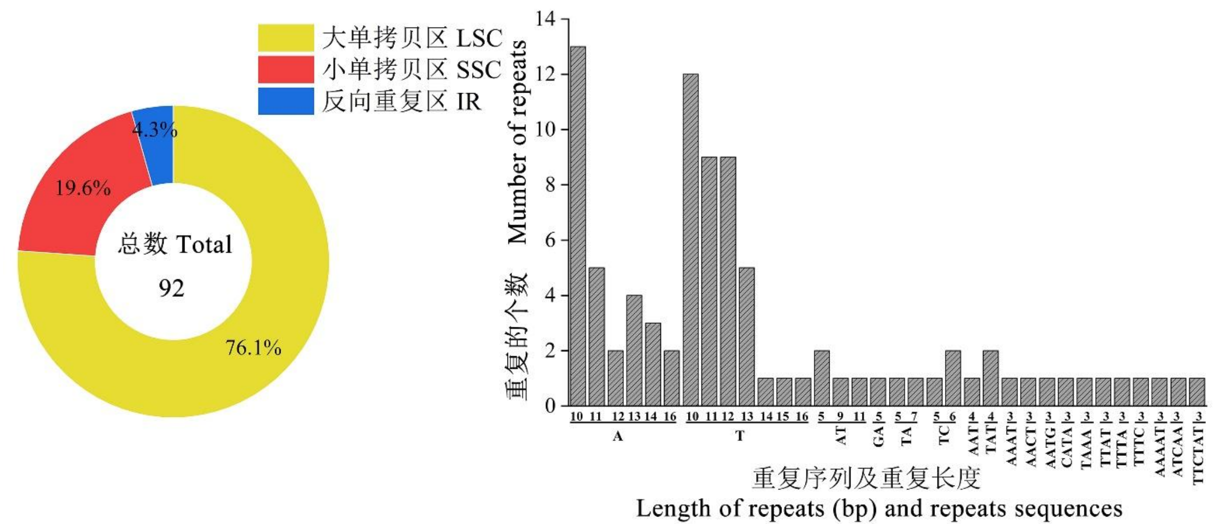


图 2 猴樟叶绿体基因组 SSR 位点分析

Fig.2 Distribution of SSR loci of *Cinnamomum bodinieri* chloroplast genome

2.3 叶绿体基因组密码子偏好性分析

猴樟叶绿体基因组有效密码子数（ENc）为 54.68，密码子适应指数（CAI）为 0.166，RSCU>1 的密码子有 30 个（表 2），其中以 A/U（T）结尾的有 26 个，以 G/C 结尾的有 4 个。鸟嘌呤和胞嘧啶碱基在密码子第三位出现频率（GC3s）为 33.8%，较叶绿体基因组 GC 含量（39.13%）低。

表 2 猴樟叶绿体基因组相对同义密码子使用情况分析

Table 2 Relative synonymous codon usage analysis of *Cinnamomum bodinieri* chloroplast genome

氨基酸 Amino acid	密码子 Codon	数量（RSCU） Number（RSCU）	氨基酸 Amino acid	密码子 Codon	数量（RSCU） Number （RSCU）
苯丙氨酸	UUU	760（1.11）*	丝氨酸	UCU	490（1.45）*
Phenylalanine	UUC	605（0.89）	Serine	UCC	379（1.12）*
亮氨酸	UUA	662（1.49）*		UCA	413（1.23）*
Leucine	UUG	578（1.30）*		UCG	209（0.62）
	CUU	513（1.15）*	脯氨酸	CCU	384（1.29）*
	CUC	265（0.60）	Proline	CCC	285（0.96）
	CUA	419（0.94）		CCA	323（1.09）*
	CUG	233（0.52）		CCG	195（0.66）
异亮氨酸	AUU	913（1.33）*	苏氨酸	ACU	437（1.39）*
Isoleucine	AUC	503（0.73）	Threonine	ACC	295（0.94）
	AUA	641（0.93）		ACA	349（1.11）*
甲硫氨酸	AUG	581（1.00）		ACG	178（0.57）
Methionine					
缬氨酸	GUU	489（1.32）*	丙氨酸	GCU	564（1.64）*
Valine	GUC	214（0.58）	Alanine	GCC	255（0.74）
	GUA	507（1.37）*		GCA	394（1.15）*
	GUG	272（0.73）		GCG	161（0.47）

酪氨酸	UAU	653（1.42）*	色氨酸	UGU	274（1.25）*
Tyrosine	UAC	266（0.58）	Tryptophan	UGC	164（0.75）
组氨酸	CAU	479（1.40）*	精氨酸	CGU	334（1.14）*
Histidine	CAC	205（0.60）	Arginine	CGC	130（0.44）
谷氨酰胺	CAA	648（1.35）*		CGA	376（1.29）*
Glutamine	CAG	309（0.65）		CGG	191（0.65）
天冬酰胺	AAU	769（1.41）*	丝氨酸	AGU	357（1.06）*
Asparagine	AAC	325（0.59）	Serine	AGC	174（0.52）
赖氨酸	AAA	911（1.38）*	精氨酸	AGA	461（1.58）*
Lysine	AAG	412（0.62）	Arginine	AGG	542（1.17）*
天冬氨酸	GAU	666（1.47）*	甘氨酸	GGU	375（0.90）
Aspartic acid	GAC	238（0.53）	Glycine	GGC	262（0.57）
谷氨酸	GAA	888（1.38）*		GGA	677（1.42）*
Glutamic acid	GAG	398（0.62）		GGG	371（0.60）
终止密码子	UAA	196（1.06）*	终止密码子	UGA	163（0.88）
Termination codon	UAG	195（1.06）*	Termination codon	UGG	443（1.00）

注: *表示 RSCU>1。

Note: * indicates RSCU>1.

2.4 IR 区边界的收缩和扩张分析

将猴樟叶绿体基因组与樟亚科内近缘属中的肉桂、檫木、新樟、月桂和大叶新木姜子叶绿体基因组进行比较，分析 IR 边界的扩张和收缩（图 3）。结果显示，6 种叶绿体基因组 IR 区域长度差异较小（20 066~20 257 bp），位于 IR 区边界及边界两侧的的基因主要包括 *ycf*2、*ycf*1、*ndh*F、*ycf*1、*ycf*2 以及 *trn*H。6 个物种 LSC-IRb（JLB）连接处的基因均为 *ycf*2 基因，SSC-IRa（JSA）连接处的基因均为 *ycf*1 基因，在 SSC 区的长度在 4 190~4 589 bp 之间，差异较大，新樟 *ycf*1 位于 IRa 区长度明显缺失。所有的 *trnH* 基因均位于 LSC 区，距 IRa-LSC（JLA）边界 1~21 bp 不等。肉桂和大叶新木姜子 IRb-SSC（JSB）边界缺失一个 *ycf*1 基因，其余 4 个物种 *ycf*1 在 938~1 430 bp 之间，长度较短，可能为假基因。新樟 JSB 边界附近缺失一个 *ndh*F 基因。

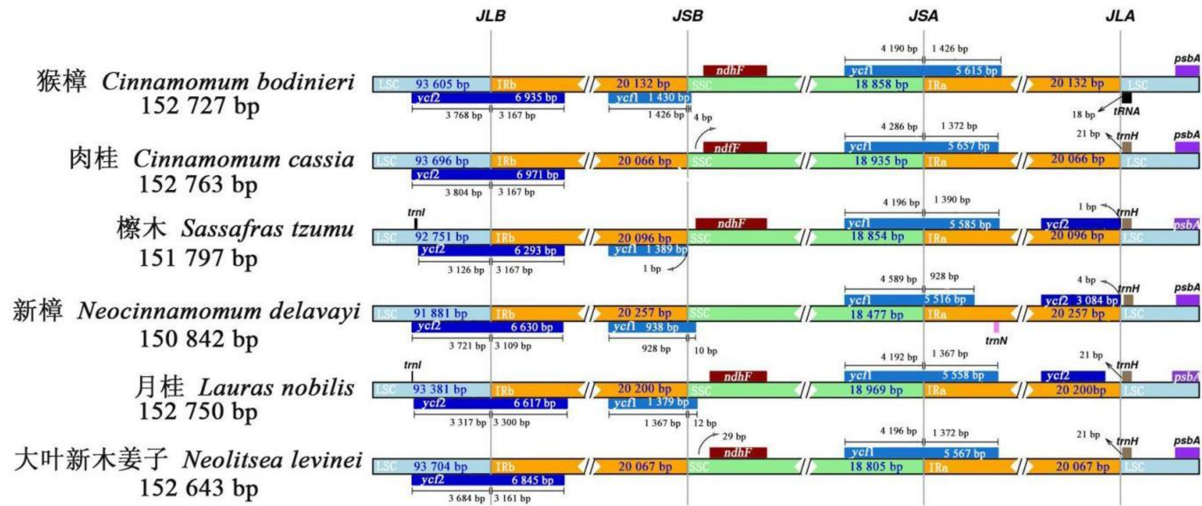


图 3 猴樟及近缘物种叶绿体基因组的 4 个连接边界

Fig.3 Four junction boundaries of the chloroplast genomes of *Cinnamomum bodinieri* and related species

2.5 系统发育分析

为明确猴樟在樟亚科的系统发育地位，利用樟亚科主要属中 24 个物种的叶绿体基因组构建了系统发育树（图 4）。根据节点支撑将构建的系统发育树分为八个分支。猴樟所在分支樟属-甜樟属分支（*Cinnamomum-Ocotea* Clade）包括樟属（*Cinnamomum*）和檫木属（*Sassafras*），该分支分为三个亚分支，樟树（*Cinnamomum camphora*）和猴樟以 100%支持率所处同一亚分支。月桂属-新木姜子属分支(*Laurus-Neolitsea* Clade)包括月桂属（*Laurus*）、木姜子属（*Litsea*）、黄肉楠属（*Actinodaphne*）、山胡椒属（*Lindera*）和新木姜子属（*Neolitsea*），其中月桂属、木姜子属和山胡椒属为复系，黄肉楠属和新木姜子属为复系。润楠属-鳄梨属分支(*Machilus-Persea* Clade)包括楠属（*Phoebe*）、鳄梨属（*Persea*）、油丹属（*Alseodaphne*）、润楠属（*Machilus*）和赛楠属（*Nothaphoebe*）。檬果樟属（*Caryodaphnopsis*）、新樟属（*Neocinnamomum*）、厚壳桂属（*Cryptocarya*）、琼楠属（*Beilschmiedia*）、油果樟属（*Syndiclis*）为单系。此外，值得注意的是，新樟属与同为樟亚族（Subtrib. Cinnamomeae）的樟属和檫木属关系较远。

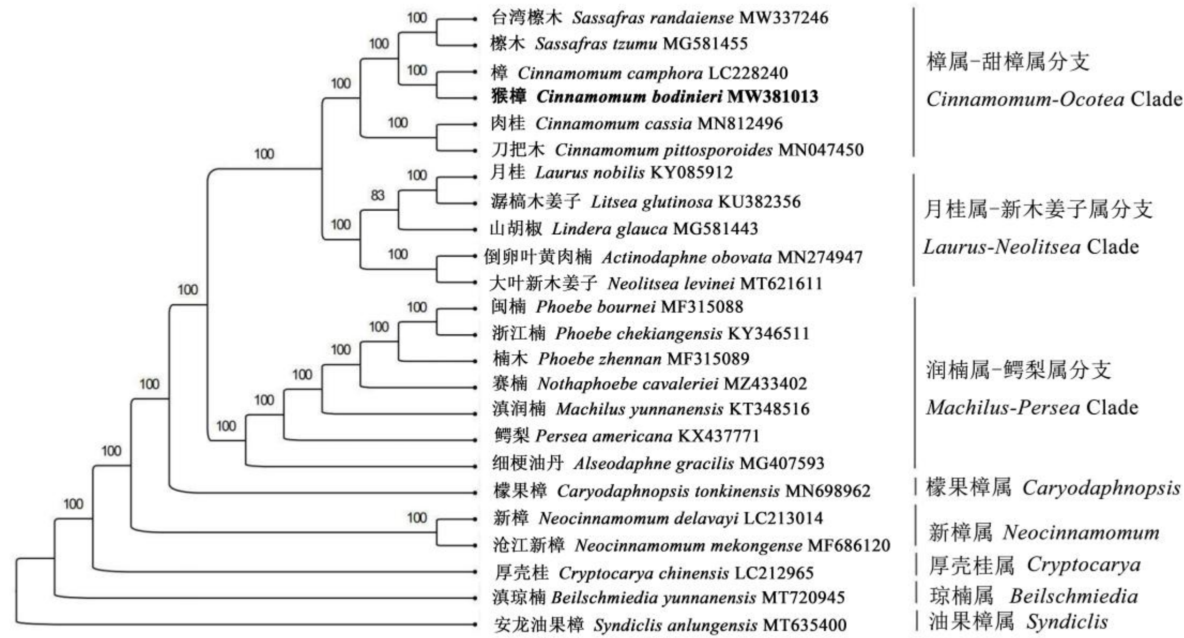


图 4 24 种樟亚科植物叶绿体基因组系统发育树

Fig.4 Phylogenetic tree of 24 species in Subfam. Lauroideae based on chloroplast genomes

3 讨论与结论

樟科植物叶绿体基因组全长在 114 603~158 598 bp 之间（田永靖，2021），差异较大，其基因组大小主要取决于 IR 区的扩张、收缩乃至缺失（Huang et al., 2014）。猴樟叶绿体基因全长 152 727 bp，IR 区全长 20 132 bp，其全长与其他樟科植物叶绿体基因组相近（Zhang et al., 2021），且 IR 区未发生明显的扩张与收缩。猴樟叶绿体基因组共编码 127 个基因，其中蛋白质编码基因 83 个，tRNA 基因 36 个，rRNA 基因 8 个，与肉桂（Song et al., 2019）、四川山胡椒（*Lindera setchuenensis*）（刘潮等，2021）、黄丹木姜子（*Litsea elongata*）（刘潮等，2022）基本一致，区别在于蛋白质编码的基因数略有不同，总 GC 含量在 39.13%~39.20%之间。由此可见，猴樟叶绿体基因组在基因组成方面与已公布的樟科植物叶绿体基因组有很大的相似性，表明在樟科植物进化过程中叶绿体基因组成相对保守。

从猴樟叶绿体基因组中搜索到 92 个 SSR 位点，其中以单核苷酸重复为主，与鼎湖钓樟（*Lindera chunii*）（Tian et al., 2019）、鳄梨（*Persea americana*）（Song et al., 2016）等其他樟科植物重复序列特征相似。猴樟叶绿体基因组 SSR 位点主要分布在单拷贝区（LSC 和 SSC），其中 LSC 区占比最

多，与猴欢喜（*Sloanea sinensis*）（王一麾等，2021）、长柄水青冈（*Fagus longipetiolata*）（Liang et al., 2021）等被子植物的研究结果一致。此外，所有的 SSR 位点中，A/T、AT/TA、AAT/TAT、AAAT/TAAA 等 A/T 重复类 SSR 占总数的 85.8%，与其他樟科植物的研究结果类似（Song et al., 2016; Tian et al., 2019）。在基因组中 GC 含量越低则 DNA 的稳定性越低，因此，较多的 A/T 重复类 SSR 可能产生更多的突变位点，这些 SSR 位点可进一步开发为樟属乃至樟科植物种内遗传变异和物种鉴定的分子标记。猴樟叶绿体基因组高频密码子有 30 个，其中 A/T（U）结尾的占 86%，这与樟树叶绿体基因密码子偏好性的研究结果相似（秦政等，2018），表明叶绿体基因组密码子第三位碱基存在不对称性，更加偏好使用 A/T（U）碱基。此外，鸟嘌呤和胞嘧啶碱基在密码子第三位出现频率（GC3s）明显较叶绿体基因组 GC 含量低，与山胡椒属（Zhao et al., 2018）、木姜子属（*Litsea*）（Zhang et al., 2021）物种研究结果类似，表明猴樟叶绿体基因组密码子更加偏好 A/T 碱基。

本研究中，6 个树种 IR 区相差较小，边界基因分布基本一致。位于 SSC、LSC 与 IR 边界的基因分别为 *ycf*1 和 *ycf*2，没有出现明显的重排现象，与田永靖（2021）对 34 种樟科植物叶绿体基因组的研究结果一致，表明樟科植物叶绿体基因组的结构具有较高的相似度，在进化上有密切的联系。被子植物中，假基因 *ycf*1（*ψycf*1）和 *ycf*2（*ψycf*2）等通常是由 IR 区的扩张与收缩产生（Gu et al., 2019），本研究中 *ψycf*1 在肉桂和大叶新木姜子中缺失，*ψycf*2 在肉桂、大叶新木姜子和猴樟中均缺失。相比其他五个物种，新樟 *ycf*1 基因出现片段缺失，*ndh*F 则完全缺失。此外，6 个物种 *ycf*1 长度差异较大，主要表现为位于 SSC 区的长度变异明显。之前的研究表明，*ycf*1 和 *ndh*F 等基因缺失和插入的频率明显高于平均水平（Daniell et al., 2016）。*ycf*1 基因对植物的生存能力至关重要，其在 SSC 区部分具有较高的序列变异性（Dong et al., 2015）；*ndh*F 基因编码叶绿体 NAD(P)H 脱氢酶(NDH)复合体的一个亚基，叶绿体 NDH 蛋白在维持植物光合作用和环境胁迫下的光保护时发挥积极作用（Peng et al., 2011）；表明 *ycf*1 和 *ndh*F 受环境选择作用较为明显，因此二者在系统发育研究中被认为是较有前途的叶绿体基因组条形码（Amar, 2020）。以上这些结果可为樟科树种亲缘关系的鉴定提供新依据。

本研究利用 NCBI 公布的樟亚科主要属的物种叶绿体基因组，通过系统进化分析确定猴樟在樟科植物中的进化地位和亲缘关系。猴樟与樟树亲缘关系最近，同时与檫木属的檫木、台湾檫木，以及樟属肉桂组的肉桂和刀把木可聚为一支，与 Zhang 等（2021）的研究结果一致。参考 Song 等（2020）基于叶绿体基因组对樟科的分类，将八个分支命名为 *Laurus-Neolitsea*、*Cinnamomum-Ocotea*、*Machilus-Persea*、*Neocinnamomum*、*Caryodaphnopsis*、*Cryptocarya*、*Beilschmiedia*、*Syndiclis* 分支。其中，樟属-甜樟属分支(*Cinnamomum-Ocotea* Clade)分支包括樟属和檫木属；月桂属、木姜子属、黄肉楠属、山胡椒属和新木姜子属归为月桂属-新木姜子属分支(*Laurus-Neolitsea* Clade)，与 Zhang 等（2021）和田永靖（2021）基于叶绿体基因组的研究结果相似。润楠属-鳄梨属分支(*Machilus-Persea* Clade)包括楠属、鳄梨属、油丹属、润楠属和赛楠属，该结果支持了 Song 等（2020）的研究。此外，Kostermans（1957）基于表型特征将樟属、檫木属、新樟属划分为樟亚族；而本研究中猴樟所在的樟属与新樟属关系较远，对 IR 边界的分析也支撑这一结果，即猴樟与新樟的 *ycf*1 和 *ndh*F 基因存在较大的差异，而这两个基因在系统发育中有重要的作用，这些结果表明新樟在进化过程中受环境选择压力作用在叶绿体基因组上表现出了较为明显的变异。Wang 等（2010）基于 A-*trnH*、*trn*K 和 *ITS* 等序列，Li 等（2016）结合 *RPB*2、*LEAFY* 和 *ITS* 序列以及田永靖（2021）基于叶绿体基因组的研究结果均表明，新樟属与樟属、檫木属分开，在分类上应该具有独立的地位，本研究所得结果也支撑了这一结论。综上所述，本文解析了猴樟叶绿体基因组的结构、基因数量、重复序列、密码子偏好等特征，分析了猴樟在樟亚科的地位以及樟亚科植物的系统发育关系，为进一步研究猴樟的系统进化和育种研究提供了理论依据。

参考文献：

AMAR MH, 2020. *ycf*1-*ndh*F genes, the most promising plastid genomic barcode, sheds light on phylogeny at low taxonomic levels in *Prunus persica* [J]. J Genet Eng Biotechnol, 18(1): 1-10.

AMIRYOUSEFI A, HYVENON J, POCZAI P, 2018. IRscope: an online program to visualize the junction

sites of chloroplast genomes [J]. Bioinformatics, 34(17): 3030-3031.
BANKEVICH A, NURK S, ANTIPOV D, et al., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing [J]. J Comput Biol, 19(5): 455-477.
BEIER S, THIEL T, MUNCH T, et al., 2017. MISA-web: A web server for microsatellite prediction [J]. Bioinformatics, 33(16): 2583-2585.
CHANDERBALI AS, WERFF H, RENNER SS, 2001. Phylogeny and historical biogeography of Lauraceae: evidence from the chloroplast and nuclear genomes [J]. Ann Mo Bot Gard, 88(1): 104-134.
COLLYDA C, DIPLARIS S, MITKAS PA, et al., 2006. Fuzzy Hidden Markov Models: a new approach in multiple sequence alignment [J]. Stud Health Technol Inform, 124: 99-104.
DANIELL H, LIN CS, YU M, et al., 2016. Chloroplast genomes: diversity, evolution, and applications in genetic engineering [J]. Genome Biol, 17(1): 1-29.
DOBROGOJSKI J, ADAMIEC M, LUCINSKI R, 2020. The chloroplast genome: A review [J]. Acta Physiol Plant, 42(1): 1–13.
DONG WP, XU C, LI CH, et al., 2015. *ycf*1, the most promising plastid DNA barcode of land plants [J]. Sci Rep, 5(1): 1-5.
FANG JY, WANG ZH, TANG ZY, 2011. Atlas of woody plants in China: distribution and climate (Volume1) [M]. Beijing and Springer-Verlag Berlin Heidelberg: Higher Education Press: 349-357.
GU CH, MA L, WU ZQ, et al., 2019. Comparative analyses of chloroplast genomes from 22 Lythraceae species: inferences for phylogenetic relationships and genome evolution within Myrtales [J]. BMC Plant Biol, 19(1): 1-19.
HUANG H, SHI C, LIU Y, et al., 2014. Thirteen *Camellia* chloroplast genome sequences determined by high-throughput sequencing: genome structure and phylogenetic relationships [J]. BMC Evol Biol, 14(1): 1-17.
HYATT D, CHEN GL, LOCASCIO PF, et al., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification [J]. BMC Bioinform, 11(1): 1-11.
KOSTERMANS AJGH, 1957. Lauraceae [J]. Reinwardtia, 4(2): 193-256.
LASLETT D, CANBACK B, 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences [J]. Nucl Acids Res, 32(1): 11-6.
LI L, MADRINAN S, LI J, 2016. Phylogeny and biogeography of *Caryodaphnopsis* (Lauraceae) inferred from low-copy nuclear gene and ITS sequences [J]. Taxon, 65(3): 433-443.
LIANG DQ, WANG HY, ZHANG J, et al., 2022. Complete Chloroplast Genome Sequence of *Fagus longipetiolata* Seemen (Fagaceae): Genome Structure, Adaptive Evolution, and Phylogenetic Relationships [J]. Life (Basel), 12(1): 92-109.
LIU C, HAN LH, PENG Y, et al., 2022. Characteristics of chloroplast genome of *Litseae longata* (Wall.ex Nees) Benth. et Hook.f. [J]. J S Agric, 53(1): 12-20. [刘潮，韩利红，彭悦，等，2022. 黄丹木姜子叶绿体基因组特征分析[J]. 南方农业学报，53(1)：12-20.]
LIU C, TANG LZ, HAN LH, 2021. Characterization of the chloroplast genome of *Lindera setchuenensis* and phylogenetics of the genus *Lindera* [J]. Sci Silv Sin, 57(12): 167-174. [刘潮，唐利洲，韩利红，2021. 四川山胡椒叶绿体基因组特征及山胡椒属系统发育[J]. 林业科学，57(12)：167-174.]
LIU ZF, MA H, CI XQ, et al., 2021. Can plastid genome sequencing be used for species identification in Lauraceae? [J]. Bot J Linn Soc, 197(1): 1-14.
NAKAMURA T, YAMADA KD, TOMII K, et al., 2018. Parallelization of MAFFT for large-scale multiple sequence alignments [J]. Bioinformatics, 34(14): 2490-2492.
NGUYEN LT, SCHIMDT HA, HAESELER A, et al., 2015. IQ-TREE: a fast and effective stochastic

algorithm for estimating maximum-likelihood phylogenies [J]. Mol Biol Evol, 32(1): 268-274.
NJUGUNA W, LISTON A, CRONN R, et al., 2013. Insights into phylogeny, sex function and age of *Fragaria* based on whole chloroplast genome sequencing [J]. Mol Phylogenet Evol, 66(1): 17-29.
PENG LW, YAMAMOTO H, SHIKANAI T, 2011. Structure and biogenesis of the chloroplast NAD(P)H dehydrogenase complex [J]. Biochim Biophys Acta, 1807(8): 945-953.
QIN Z, ZHENG YJ, GUI LJ, et al., 2018. Codon usage bias analysis of chloroplast genome of camphora tree (*Cinnamomum camphora*) [J]. Guihaia, 38(10): 1346-1355. [秦政，郑永杰，桂丽静，等，2018. 樟树叶绿体基因组密码子偏好性分析 [J]. 广西植物，38(10)：1346-1355.]
SHINOZAKI K, OHME M, TANAKA M, et al., 1986. The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression [J]. EMBO J, 5(9): 2043-2049.
SONG FL, ZHANG MM, SU JL, et al., 2012. The comparison on cold resistance between *Cinnamomum bodinieri* and *Cinnamomum camphora* seedlings in natural decreasing process of air temperature [J]. J West Chin For Sci, 41(6): 48-52. [宋芳琳，张苗苗，苏金乐，等，2012. 自然降温过程中猴樟和香樟幼苗的抗寒性比较 [J]. 西部林业科学，41(6)：48-52.]
SONG K, HE M, YU JB, et al., 2019. Characterization of the chloroplast genome of the family Lauraceae plant species, *Cinnamomum cassia* [J]. Mitochondrial DNA Part B, 4(2): 3906-3907.
SONG Y, YAO X, TAN YH, et al., 2016. Complete chloroplast genome sequence of the avocado: gene organization, comparative analysis, and phylogenetic relationships with other Lauraceae [J]. Can J For Res, 46(11): 1293-1301.
SONG Y, YU WB, TAN YH, et al., 2020. Plastid phylogenomics improve phylogenetic resolution in the Lauraceae [J]. J Syst Evol, 58(4): 423-439.
TIAN XY, YE JW, SONG Y, 2019. Plastome sequences help to improve the systematic position of trinerved *Lindera* species in the family Lauraceae [J]. PeerJ, 7: e7662.
TIAN XQ, WEI XL, 2011. Effect of different ecological environments on plantation growth of *Cinnamomum bodinieri* in Guizhou [J]. Guizhou Agric Sci, 39(4): 184-187. [田小琴，韦小丽，2011. 生态环境对贵州猴樟人工林生长的影响 [J]. 贵州农业科学，39(4)：184-187.]
TIAN YJ, 2021. The complete chloroplast genomes of Lauraceae species: comparative geomic and phylogenetic analyses [D]. Nanjing: Nanjing University: 5-7. [田永靖，2021. 樟科植物比较叶绿体基因组与系统发育研究 [D]. 南京：南京大学：5-7.]
WANG YH, XIE YF, ZHANG ZH, et al., 2021. The complete chloroplast genome of *Sloanea sinensis* and the systematic status of Eloeacarpaceae [J]. Guihaia, 42(1): 39-48. [王一麾，谢宜飞，张志翔，等，2021. 猴欢喜叶绿体全基因组及杜英科系统地位分析 [J]. 广西植物，42(1)：39-48.]
WANG ZH, LI J, CONRAN JG, et al., 2010. Phylogeny of the southeast Asian endemic genus *Neocinnamomum* H. Liu (Lauraceae) [J]. Plant Syst Evol, 290(1): 173-184.
WONG EHM, SMITH DK, RABADAN R, et al., 2010. Codon usage bias and the evolution of influenza a viruses. Codon Usage Biases of Influenza Virus [J]. BMC Evol Biol, 10(1): 1-14.
XIAO ZF, JIN ZN, ZHANG BH, et al., 2020. Effects of IBA on rooting ability of *Cinnamomum bodinieri* citral type micro-shoots from transcriptomics analysis [J]. Plant Biotechnol Rep, 14(4): 467-477.
XIAO ZF, WANG LL, CAO LY, et al., 2020. Tissue culture technology of stem segment of *Cinnamomum bodinieri* var. *citralifera* [J]. Bull Bot Res, 40(2): 196-201. [肖祖飞，王玲玲，曹璐瑶，等，2020. 柠檬醛猴樟茎段组织培养技术研究 [J]. 植物研究，40(2)：196-201.]
ZHANG YY, TIAN YJ, TNG DYP, et al., 2021. Comparative chloroplast genomics of *Litsea* Lam. (Lauraceae) and its phylogenetic implications [J]. Forests, 12(6): 744-758.
ZHANG Y, WEI XL, WANG L, et al., 2014. Growth variability of *Cinnamomum bodinieri* seedlings from

different geographical provenances [J]. SW Chin J Agric Sci, 27(5): 2162-2167. [张怡，韦小丽，王娈，等，2014. 不同地理种源猴樟苗期生长变异性 [J]. 西南农业学报，27(5)：2162-2167.]
ZHAO ML, SONG Y, NI J, et al., 2018. Comparative chloroplast genomics and phylogenetics of nine *Lindera* species (Lauraceae) [J]. Sci Rep, 8(1): 1-11.